Symbolic Music Generation with Diffusion Models

Year: 2021 | Citations: 213 | Authors: Gautam Mittal, Jesse Engel, Curtis Hawthorne, Ian Simon

---

Abstract

Score-based generative models and diffusion probabilistic models have been successful at generating high-quality samples in continuous domains such as images and audio. However, due to their Langevin-inspired sampling mechanisms, their application to discrete and sequential data has been limited. In this work, we present a technique for training diffusion models on sequential data by parameterizing the discrete domain in the continuous latent space of a pre-trained variational autoencoder. Our method is non-autoregressive and learns to generate sequences of latent embeddings through the reverse process and offers parallel generation with a constant number of iterative refinement steps. We apply this technique to modeling symbolic music and show strong unconditional generation and post-hoc conditional infilling results compared to autoregressive language models operating over the same continuous embeddings.